# 轉譯生醫資訊學 期末報告說明

## 情境說明

假設本所配合教育部「建構大學衍生新創研發服務公司之孕育機制」政策，開設**揚名國際公司**，提供研究服務，協助研究者收集、整合、管理大數據，走向利用人工智慧(AI)改善人類健康的新時代。如圖1所示，公司具有研發部與產品部，兩者相互配合發展人工智慧所衍生的產品。在這兩個單位中，共有5個不同的小組，分別是

1. **數據中心**: 利用行動裝置APP ，協助使用者收案、清理數據、典藏數據、釋出數據、更新數據版本、回存子數據集等。
2. **分析中心**: 分析數據中心典藏的數據，產生關連性，或訓練出AI模型。
3. **影像核心設施**: 協助數據中心與分析中心處理(典藏或分析)影像相關的數據
4. **行動裝置APP**: 將分析中心訓練出的模型佈署到行動裝置的APP上，提供使用者服務，並協助影像註解
5. **知識庫**: 將分析中心得到的關聯性，或是衛教資訊，佈署到知識庫中，再利用LINE bot提供使用者服務

公司現階段以發展腦中風的健康保健為主軸，因此以「暫時性腦缺血(transient ischemic attack, TIA)」的病人為對象發展產品。因為這類型的病人再次得到腦中風的風險相當高，所以比較需要提供衛教資料給病人，讓他們可以預防二次腦中風。揚名國際公司製作腦中風APP，協助暫時性腦缺血的病人評估再次得到腦中風的風險，並提供各種衛教資料，教導病人如何避免再次得到腦中風。此外也提供測試工具，讓病人覺得不適時，可以判斷自己是否得到第二次的腦中風，以利即時就醫。

透過產學合作，醫院急診室在發現有暫時性腦缺血的病人時，會通知揚名國際公司的志工，到場協助病人安裝腦中風APP，並解釋使用的方式。同時也會鼓勵病人捐贈自己的健康資料，以協助研究預防腦中風的方法，避免其他人得到腦中風。若病人願意捐贈自己的健康數據，則以謹慎和嚴謹的方式，使用這些捐贈者提供的資料，進一步的開發產品，造福人類。

## 課程期末報告說明

### 數據中心的網站(由課程助教提供)

網站的公開網頁會有簡介部份說明數據釋出政策、數據清理原則等；有帳號申請部份，串聯到追蹤的流程；有使用者常問問題(FAQ)的部份，讓使用者求助；甚至提供數據探索的功能。

申請帳號時**，**會依據據釋出政策所描述的申請者類別，決定使用者可以看到哪些數據集。使用者一旦申請到帳號，即可瀏覽自己權限內可看到的數據集簡述。在簡述頁的列表中會顯示可下載的編碼簿與釋出文件，使用者可以根據這些文件決定要下載哪一個數據集，在按下「下載」鈕時，系統會要求使用者簽署保密協定，簽完後才能取得數據。簽完保密協定後，即可取得屬於同一個數據集的子數據集，例如回存的數據集。

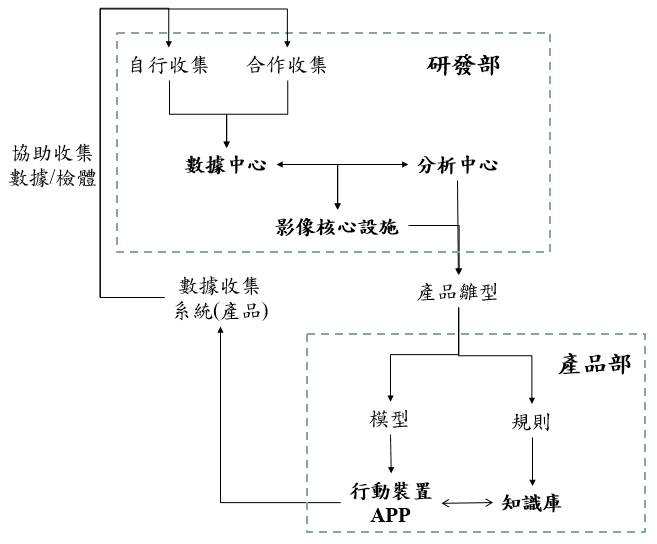


圖1 揚名國際公司產品開發流程

以下將每一組的工作，做一個比較仔細的說明

### 行動裝置APP組

除了開發出腦中風APP外，也發展了嚴謹的電子化知情同意流程，讓病人可以充分的瞭解捐贈數據的意義，與臨床研究進行的方式。病人在了解狀況的情形之下，可決定是否要捐贈自己的健康資料。這些數據需要存成AI-ready的形式，方便未來做人工智慧研究時取用這些資料，因此數據將直接存入符合國際資料交換標準的FHIR伺服器。

### 知識庫組

收集了許多衛教資料，包括教導正常人如何預防腦中風，或教導腦中風病人如何預防二次中風等。他們將這一些衛教資料分類，並繪製出簡單的圖像引導使用者，選擇到適當的衛教資料。此外，也設計聊天機器人(chatbot)的技術，利用病人所捐贈的病史資料和健康狀態，甚至天候等環境因子，與分析組發現的規則，決定如何在最恰當的時間，推送最恰當的衛教資料給病人。

### 數據中心組

可以由FHIR伺服器中，讀取病人捐贈的資料。數據中心組的工作目標就是要將這些原始數據經過嚴謹的整理與清理過程，得到可做後續分析的數據。處理數據的目的是希望說數據可以和其他系統交換(interoperatable)，以增加系統的再使用性(reusable)。在處理數據的過程中必須有嚴謹的去識別化和版本控管過程，提升數據的可近性(accessible)。並建立中心網站，讓使用者可以輕易找到(findable)數據，希望符合FAIR的原則。此外，也要建立數據回存的機制，讓分析過的數據能存回數據中心，以達到其他的研究者可再現(reproducible)已發表研究結果的目的。這些回存的數據，也可以節省其他的分析者清理數據所需花的時間，因此當資料有更新時，也需要利用程式更新這些回存的數據。

### 影像核心設施

影像數據是現在人工智慧研究經常使用的數據，他們的管理方式和一般的文字行數據不同因此特別設立影像核心設施一方面協助管理影像數據另一方面也協助使用者來使用這些影像數據。取得影像時也必須要做去識別化，儲存的數據有時需要做註解，然後再回存，方便未來做AI研究時可以知道正確答案是什麼。使用者的帳號與數據中心網站分別控管，但是可以透過數據中心的管理機制，申請帳號或申請數據的使用。有關影像與影像報告入庫所需要用的程式，也會做版本控管。

### 分析中心

分析中心需要由數據中心或影像核心設施取得數據，做後續的分析。分析通常分為兩個階段，第一個階段是探索式的分析，希望找到值得繼續研究的題目。一旦找到適合的主題，可能需要對於選定的欄位進行邏輯性的清理，取得清理好的數據後再做深入的分析。因為第一階段已有初步的觀察結果，第二個階段通常需要驗證假說是否正確。通常對於建立的模型，需要做性能的評估，例如找出最恰當的運作條件，並計算靈敏度、專一性、正確性、精準度等，以顯示模型的優越性能。

## 期末報告順序與內容

### 第四組 APP組

#### 示範如何協助TIA患者設定APP

首先要做知情同意，取得同意後協助患者照相、錄音，接下來進行中風的風險評估，並教導患者如何做FAST評估。再教導患者如何使用衛教資料，並設定連到「衛教小幫手」，以得到互動式的衛教資訊。最後協助患者捐贈健康存摺的數據。示範步驟如下:

1. 利用電子化的知情同意(eConsent)工具，向患者說明腦中風研究可以造福其他人，參與研究時需要做哪些步驟，會如何保護他的個資等。如果患者願意參與，則簽署知情同意書，並協助他下載健康存摺數據。不論患者是否願意參與，均協助患者進行下列步驟:
2. 協助TIA患者安裝腦中風App
3. 協助TIA患者進行腦中風的風險分析(這裡包含兩種方法，第一種是Eisen所發展的方法，第二種是揚名國際公司所建立的模型)
4. 示範如何取得衛教資料讓TIA患者了解如何預防腦中風的發生
5. 示範如何輸入GUID，訂閱腦中風的衛教小幫手，方便產生互動式的衛教
6. 示範如何使用FAST測試，確認自己是否發生第二次的腦中風
7. 連線到FHIR伺服器，顯示知情同意、照相、錄音、評估風險時所記錄的疾病史、與患者捐贈的健康存摺等數據，都已放到FHIR伺服器
8. 展示Gitlab上整理的程式與「說明文件」。

### 第五組 知識庫組

#### 展示「衛教小幫手」的瀏覽功能

首先說明如何註冊使用「衛教小幫手」，讓聽眾可以一起測試並提供意見。接下來示範「衛教小幫手」的功能，例如教中風患者瞭解預後，學習如何照顧自己避免惡化、如何復健等；再如教健康人如何保健，降低得到腦中風的風險。

#### 展示「衛教小幫手」的推送功能

首先帶大家瀏覽所推送文章的主要內容，然後再查閱患者在FHIR伺服器上登錄的疾病史資料，以顯示「衛教小幫手」能夠依據患者的狀況，推送適當的衛教文章。

### 第一組 數據中心

#### 利用自行收錄資料，示範具有AI-ready的功能

(利用美兆健康資料庫的數據，模擬自行蒐集的資料，在FHIR伺服器的資料庫中準備可供展示的大量數據。其中有一些有資料缺失，另一些有填錯的情形，讓我們可以展示資料清理的效果)

1. 首先利用投影片說明工作流程: 針對自行蒐集的數據內容，設計清理的流程，並依據可以取得的數據格式，撰寫自動進行數據清理的Web service程式。再決定要呼叫哪些API程式，以取得FHIR伺服器內的數據。最後撰寫腳本程式，串接取得自行收錄資料的，和數據清理所使用的API程式，將自行收集的數據整理成可以進行分析的csv格式。
2. 其次展示FHIR伺服器內的數據內容，指出資料缺失，與填錯的部份。再執行程式並呈現經過程式處理後的結果，顯示數據已處於AI-ready的狀態。

#### 示範數據中心的運作與管理方式

1. 先檢視是否有知情同意，若存在才可取用數據
2. 展示如何利用程式到FHIR伺服器取得原始數據，並且為檔案命名，編上版號
3. 展示如何檢視原始數據，找出可能的錯誤並且定義什麼是outlier
4. 展示如何制定清理的規則，據此撰寫程式，並做版本控管(在Gitlab上展示版號)
5. 展示如何製作編碼簿(codebook)，撰寫釋出文件(release note) - 請特別注意版本控管的問題
6. 展示如何在網站上放置可以釋出的數據與文件，並編寫網頁讓使用者可以找到這筆新的數據
7. 展示數據回存時，如何將數據、文件編版號，將程式上傳到Gitlab，並將數據在中心的網站上架
8. 展示數據更新時，如何準備另一筆新版的數據集
9. (展示數據的版本更新時，如何更新回存的數據)

### 第三組 影像中心 Xnat

1. 將數據去識別化
2. 影像入庫
3. 影像報告入庫
4. 查詢
5. 註解並回存影像
6. 程式版本控管

### 第二組 分析中心

1. 在數據中心的網站上申請帳號
2. 瀏覽網站上的數據集，由編碼簿、釋出文件等，選擇需要使用的數據集
3. 簽署保密協定，下載數據
4. 選擇要分析的數據，並作邏輯性清理
5. 分析數據，建立模型，並對程式做版本控管
6. 將模型與測試數據交給APP組
7. 準備將清理好的子資料集回存到數據中心，包括程式編碼簿釋出文件等